



### **Hak cipta dan penggunaan kembali:**

Lisensi ini mengizinkan setiap orang untuk mengubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

### **Copyright and reuse:**

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

# **BAB I**

## **PENDAHULUAN**

### **1.1. Latar Belakang Masalah**

*Cyberbullying* merupakan segala bentuk penggunaan Internet, telepon selular, atau peralatan lainnya yang ditujukan melukai ataupun mempermalukan orang lain dengan cara mengirimkan teks atau gambar kepada orang tersebut (Agrawal and Awekar, 2018). 49% dari 5900 sampel pengguna Internet di Indonesia pernah mengalami *cyberbullying* pada *platform* media sosial (Pratomo, 2019). Beberapa dampak dari tindakan *cyberbullying* yang dirasakan oleh korban adalah munculnya ketakutan, menurunnya tingkat kepercayaan diri, depresi, insomnia, dan bahkan kematian yang diakibatkan oleh kasus bunuh diri (Tamtomo, 2019). Pihak media sosial sudah menyediakan sebuah layanan yang dapat digunakan oleh penggunanya untuk melaporkan adanya tindakan *cyberbullying*, tetapi hal ini masih belum efektif untuk dijalankan karena pelaksanaan tindak lanjut suatu laporan membutuhkan waktu sehari-hari (BBC Technology, 2018). Berdasarkan permasalahan di atas, adanya deteksi ujaran atau komentar *cyberbullying* menjadi penting karena dapat membantu penanganan konten *cyberbullying* yang tersebar pada *platform* media sosial menjadi lebih baik dan lebih efisien (Purnamasari et al., 2018).

Menurut penelitian sebelumnya yang dilakukan oleh Andriansyah, dkk (2017), algoritma *Support Vector Machine* dapat menghasilkan nilai *accuracy* sebesar 79,412% dalam melakukan klasifikasi komentar *cyberbullying*. Algoritma ini berhasil melakukan klasifikasi 27 komentar dari 34 data komentar dengan tepat. Algoritma *Support Vector Machine* juga digunakan dalam penelitian lainnya, yaitu

penelitian Darson (2019) yang menggunakan TF-IDF dan *Support Vector Machine* untuk melakukan klasifikasi komentar *cyberbullying*. Penelitian tersebut mendapatkan nilai *accuracy* sebesar 76,77%, *precision* sebesar 78,88%, *recall* sebesar 74,60%, dan *f-measure* sebesar 76,68%.

Menurut Sahrul, dkk (2019), algoritma *Support Vector Machine* tidak membantu komputer dalam memahami konteks suatu ujaran, sehingga algoritma ini dianggap kurang efektif dalam melakukan klasifikasi teks. Oleh karena itu, algoritma yang digunakan dalam penelitian tersebut adalah *Artificial Neural Network* (ANN) dan *Recurrent Neural Network* (RNN). Berdasarkan penelitian tersebut, didapatkan hasil dari algoritma RNN lebih baik dibandingkan dengan ANN. RNN menghasilkan nilai *true positive* dan *true negative* dari *confusion matrix* sebesar 83,8% dan 84,4%, sedangkan ANN menghasilkan nilai *true positive* dan *true negative* sebesar 82,2% dan 80,6%.

Anindyati, Purwarianti and Nursanti (2019) melakukan penelitian untuk mendeteksi teks *cyberbullying* dengan menggunakan tiga algoritma, yaitu algoritma Long Short-Term Memory (LSTM) yang merupakan pengembangan dari algoritma RNN, Bidirectional Long Short-Term Memory (Bi-LSTM) yang merupakan pengembangan dari algoritma LSTM, dan Convolutional Neural Network (CNN). Selain tiga algoritma tersebut, penelitian ini menggunakan pemodelan Word2Vec sebagai *word embedding*. Penelitian tersebut mendapatkan nilai *accuracy* sebesar 90,57% dan nilai *recall* untuk kelas *bullying* sebesar 75,7%. Hasil tersebut didapatkan dengan menggunakan algoritma LSTM karena algoritma LSTM dapat menghasilkan nilai evaluasi tertinggi dibandingkan dengan nilai evaluasi yang dihasilkan oleh algoritma Bi-LSTM dan CNN.

Penelitian deteksi kalimat umpatan yang dilakukan oleh Ibrohim, Sazany, dan Budi (2019) menggunakan algoritma LSTM. Selain itu, penelitian ini juga menggunakan *word embedding* untuk merepresentasikan kata-kata ke dalam vektor. *Word embedding* yang digunakan, yaitu Word2Vec dan Fasttext. Penelitian tersebut mendapatkan nilai  $F_1$ -score sebesar 83,68%. Hasil tersebut didapatkan dengan menggunakan Fasttext karena Fasttext dapat membantu algoritma LSTM melakukan deteksi kalimat umpatan lebih baik dibandingkan dengan menggunakan Word2Vec. Hasil ini lebih baik dibandingkan dengan penelitian Ibrohim dan Budi (2018) yang menggunakan beberapa algoritma, termasuk *Support Vector Machine*. Nilai  $F_1$ -score yang didapatkan dengan menggunakan *Support Vector Machine* sebesar 67,48%.

Sazany dan Budi (2018) juga menggunakan algoritma LSTM dalam melakukan penelitian untuk mengidentifikasi *hate speech* dalam teks. Selain LSTM, penelitian menggunakan Word2Vec dan Fasttext dalam melakukan *word embedding*. Nilai  $F_1$ -score yang didapat dari penelitian tersebut sebesar 95,93% dengan menggunakan Fasttext sebagai *word embedding* karena hasil tersebut lebih baik dibandingkan dengan menggunakan Word2Vec sebagai *word embedding*. Hasil ini juga lebih baik dibandingkan dengan penelitian Alfina dkk. (2017). Penelitian tersebut menggunakan beberapa algoritma, termasuk *Support Vector Machine*. Nilai  $F_1$ -score yang didapatkan dengan menggunakan *Support Vector Machine* sebesar 72,3%.

Berdasarkan penelitian-penelitian terkait yang sudah dijelaskan, diketahui bahwa *word embedding* juga dapat digunakan untuk membantu sistem deteksi menghasilkan hasil deteksi yang akurat. *Word embedding* merupakan teknik yang

digunakan untuk merepresentasikan kata-kata ke dalam vektor. Teknik ini dilakukan karena komputer menerima data berupa vektor sebagai *input* (Tensorflow, 2020). Terdapat dua model *word embedding* yang dapat digunakan berdasarkan penelitian-penelitian terkait, yaitu Word2Vec dan Fasttext. Word2Vec merupakan pemodelan *word embedding* yang menggunakan pendekatan *dictionary* dalam merepresentasikan kata-kata. Hal ini mengakibatkan Word2Vec tidak dapat merepresentasikan kata-kata yang belum pernah dipelajari sebelumnya, sehingga terdapat kata-kata yang tidak memiliki representasi vektor (Choi and Lee, 2020). Fasttext dikembangkan untuk mengatasi permasalahan tersebut dengan menggunakan pendekatan *syntactic* atau struktur kata. Hal ini dapat dilakukan karena Fasttext menggunakan *character n-gram*. Metode *character n-gram* memungkinkan Fasttext untuk merepresentasikan kata-kata ke dalam vektor, bahkan dapat merepresentasikan kata-kata yang belum dipelajari sebelumnya (Choi and Lee, 2020).

Berdasarkan penjelasan latar belakang yang sudah dijelaskan, diketahui bahwa algoritma Long Short-Term Memory dapat diimplementasikan untuk melakukan deteksi ujaran atau komentar *cyberbullying*. Algoritma LSTM dipilih karena mampu memberikan performa yang lebih baik dibandingkan algoritma-algoritma lain yang digunakan dalam penelitian-penelitian terkait. Selain algoritma LSTM, Fasttext juga diimplementasikan dalam sistem sebagai *word embedding* karena mampu merepresentasikan kata-kata yang bahkan belum dipelajari sebelumnya yang tidak mampu dilakukan oleh Word2Vec. Selain itu, penelitian terkait menunjukkan bahwa nilai evaluasi yang didapatkan dengan menggunakan Fasttext sebagai *word embedding* lebih baik dibandingkan dengan Word2Vec.

## 1.2. Rumusan Masalah

Berdasarkan latar belakang yang sudah dibahas, ditentukan rumusan masalah sebagai berikut.

1. Bagaimana mengimplementasikan algoritma *Long Short-Term Memory* untuk mendeteksi komentar *cyberbullying* dalam bahasa Indonesia?
2. Berapa nilai *precision*, *recall*, dan *f-measure* dari algoritma *Long Short-Term Memory* dalam mendeteksi komentar *cyberbullying* dalam bahasa Indonesia?

## 1.3. Batasan Masalah

Batasan masalah yang digunakan pada penelitian ini adalah sebagai berikut.

1. Penelitian ini akan membedakan komentar menjadi dua kelas, yaitu *cyberbullying* dan bukan *cyberbullying*.
2. Penelitian ini menggunakan *dataset* yang digunakan dari penelitian yang dilakukan oleh Andriansyah, dkk (2017) untuk melakukan *training* algoritma LSTM.
3. Penelitian ini menggunakan *dataset* yang berisi 1052 data yang terdiri atas 435 data dengan label *cyberbullying* (1) dan 617 data dengan label bukan *cyberbullying* (0).

## 1.4. Tujuan Penelitian

Penelitian ini dilakukan dengan tujuan untuk:

1. Mengimplementasikan algoritma *Long Short-Term Memory* untuk mendeteksi komentar *cyberbullying* dalam bahasa Indonesia.

2. Mengetahui nilai *precision*, *recall*, dan *f-measure* dari algoritma *Long Short-Term Memory* dalam mendeteksi komentar *cyberbullying* dalam bahasa Indonesia.

### **1.5. Manfaat Penelitian**

Manfaat yang dapat diberikan melalui penelitian ini adalah mengembangkan suatu metode yang dapat melakukan deteksi suatu komentar atau ujaran *cyberbullying* dengan akurat.